

CLAIMS

1 1. A method for load balancing a plurality of servers, the method comprising:
2 providing a plurality of control blocks, each control block associated with a num-
3 ber of active connections a server is connected with, the control block configured to con-
4 trol at least one server with the associated number of connections in a server list;
5 causing each control block to point to a server with a least value ascertained by
6 determining the number of connections on the server relative to the server's capacity to
7 handle connections;
8 selecting the control block associated with the least number of connections; and
9 selecting the server pointed to by the control block.

1 2. The method as in claim 1, wherein ascertaining the least value for the
2 server comprises:
3 determining a metric of the server by dividing the number of connections on the
4 server by the capacity of the server, wherein the metric is kept as a quotient/remainder
5 pair;
6 storing the quotient/remainder pair in the control block;
7 incrementing the remainder by one for every connection added to the server; and
8 decrementing the remainder by one for every connection removed from the server.

1 3. The method as in claim 1, further comprising:
2 causing the control block with the server having an added/removed connection to
3 transfer the server to an adjacent control block, wherein the adjacent control block is as-
4 sociated with the number of connections pertaining to the transferring server;
5 causing the control block to transfer the metric of the server to the adjacent con-
6 trol block; and
7 updating the pointer to point to the next server on the list of the control block.

1 4. The method as in claim 3, further comprising:
2 removing the control block if the control block does not have a server on the server list.

- 1
- 2
- 3
- 4
- 5

- 1
- 2
- 3

1
2
3
4
5
6
7
8
9
10

1
2
3
4
5
6
7
8

1 14. The load balancing apparatus as in claim 13, further comprising:
2 The control block configured to determine a metric of the server, wherein the met-
3 ric is kept as a quotient/remainder pair;
4 a memory to store the quotient/remainder pair determined by the control block;
5 the control block further configured to increment the remainder by one for every
6 connection added to the server; and
7 the control block further configured to decrement the remainder by one for every con-
8 nection removed from the server.

1 15. The load balancing apparatus as in claim 13, further comprising:
2 the control block configured to transfer the server having an added/removed con-
3 nection to an adjacent control block, wherein the adjacent control block is associated with
4 the number of connections pertaining to the transferring server;
5 the control block further configured to transfer the metric of the server to the ad-
6 jacent control block; and
7 the control block configured to update the pointer to point to the next server on the list of
8 the control block.

1 16. The load balancing apparatus as in claim 15 further comprises:
2 the control block is de-activated if the control block does not have a server on the server
3 list.

1 17. The load balancing apparatus as in claim 15, further comprises:
2 the adjacent control block configured to receive the transferring server; and
3 the adjacent control block further configured to receive the metric of the transferring
4 server, wherein the adjacent control block updates and sorts the server list.

1 18. The load balancing apparatus as in claim 17, further comprises:

- 17